Journal of Multivariate Analysis **75**, 47–61 (2000)
doi:10.1006/jmva.1999.1893, available online at http://www.idealibrary.com on IDEAL®

Fitting Smooth Histories to Rotation Data[1]

Martin S. Hanna

Department of Mathematics, University of Kansas, Lawrence, Kansas 66045
E-mail: hanna@math.ukans.edu

and

Ted Chang

Division of Statistics, University of Virginia, Charlottesville, Virginia 22903
E-mail: tcc8v@virginia.edu

Received September 10, 1998; published online August 31, 2000

Consider two tectonic plates diverging at a mid-ocean ridge. Geophysicists are able to estimate the rotation of one plate relative to the other at a discrete sequence of times in the earth's history; also they usually have information as to the likely errors in these rotation estimates. We address the problem of fitting a smooth history to such rotation data. We employ a modification of the method used by Jupp and Kent in their 1987 article dealing with fitting a smooth history to time-labeled points on the surface of the unit sphere in three-dimensional space. They use parallel translation to "unroll" data from the surface of the sphere to a plane. We replace unrolling via parallel translation by unrolling via left group multiplication, using the group structure of SO(3). We explain why our understanding of the errors in tectonic plate reconstructions dictates that left group multiplication is preferable both to parallel translation and to right group multiplication. To choose the smoothing parameter we use the discrepancy method; for the Central Atlantic data set which we consider this method gives considerably better results than cross-validation.

AMS 1991 subject classifications: 62.

Key words and phrases: plate tectonics, rotations on spheres, smoothing spline.

1. INTRODUCTION

In the study of tectonic plate motion, researchers are able to estimate the rotation of one plate relative to another at a discrete sequence of times (the

[1] The authors thank Professors Jupp and Kent for access to the computer program described in their article [9]. This work was partially supported by National Science Foundation Grant DMS-9404585 and Office of Naval Research Grant N00014-92-J-1009. Copies of FORTRAN programs implementing the procedures outlined in this article will be available at http://www.stat.virginia.edu.

0047-259X/00 $35.00

times of the reversals of the earth's magnetic field). In addition, it is common that information is available as to the likely errors in these rotation estimates. The problem which we address in this article is that of fitting a smooth history to such rotation data, so that estimates can be made of the rotations obtaining at times between the times where estimates are already available.

To be specific, let $\tau_1 < \tau_2 < \cdots < \tau_n$ be the discrete sequence of times, where τ_i represents a time before the present measured in millions of years. Let A_i be the true rotation corresponding to τ_i and let $\hat{A}_i$ be the estimate for A_i which is available. To the data $(\tau_i, \hat{A}_i)$ we seek to fit a smooth history $\hat{A}(t)$, so that $\hat{A}(t)$ constitutes an estimate of the true rotation $A(t)$ at time t.

In order to perform the fitting discussed, it is necessary to parametrize the rotations A of $\mathbb{R}^3$. We use the parametrization in terms of unit quaternions. Namely, we use the mapping φ of $\mathbb{Q}$, the set of unit quaternions, onto SO(3), the group of rotations of $\mathbb{R}^3$, defined as follows: for $q \in \mathbb{Q}$, $A = \varphi(q)$ means

$$Ax = qx\bar{q},$$

where x is a pure quaternion. (We identify the pure quaternions with column vectors in $\mathbb{R}^3$. The product on the left above is a matrix product, while the products on the right are quaternion products.) Note that if q is written in the form

$$q = \cos\left(\frac{\rho}{2}\right) + \sin\left(\frac{\rho}{2}\right) u,$$

where u is a unit-length pure quaternion and $0 \leqq \rho \leqq 2\pi$, then $A = \varphi(q)$ is the right-hand rotation of ρ radians about the axis u. The representation of rotations in terms of unit quaternions has many other desirable properties; in particular, the mapping φ is essentially distance-preserving (see Chang *et al.* [5]).

Let w_i be the unit quaternion corresponding to the rotation $\hat{A}_i$. Our data then consist of a sequence of time-labeled points (τ_i, w_i), where $w_i \in S^3$, the unit sphere in $\mathbb{R}^4$. (We identify $\mathbb{Q}$ with S^3.)

We describe the likely errors in the estimate $\hat{A}_i$ as follows. An asymptotic 95% confidence region for the true rotation A_i consists of all rotations $A = \varphi(q)$ where

$$q^t \tilde{Q}_i q \leqq 1.$$

Here $\tilde{Q}_i$ is a 4×4 positive semi-definite matrix; it has three positive eigenvalues, plus one zero eigenvalue corresponding to the eigenvector w_i. The

description of confidence regions in this form is discussed in Hanna and Chang [6, pp. 167–168, 174]. The derivation of such confidence regions, under a number of different probability assumptions, is given in Chang [1–4] and Rivest [11].

2. PREVIOUS WORK

The fitting problem discussed above was considered by Prentice [10]. Prentice's procedure is to first rotate S^3 so that the data points w_i are grouped as closely as possible about the pole $1 = (1, 0, 0, 0)^t$. Prentice then projects the rotated data points onto $T_1 S^3$, the tangent space to S^3 at 1 (which we can think of as $\mathbb{R}^3$), using the inverse exponential map.

Note that the exponential map, which we denote by "exp," maps a vector $x \in T_1 S^3$ onto a point $r = \exp(x)$ on S^3 according to

$$r = \exp(x) = \begin{cases} \cos(|x|) + \sin(|x|)\, x/|x|, & \text{if } x \neq 0 \\ 1 & \text{if } x = 0. \end{cases}$$

The exponential map transforms the open ball $|x| < \pi$ in $T_1 S^3$ in a one-to-one manner onto $S^3 - \{-1\}$. Note also that the spherical distance from $\exp(x)$ to 1 is $|x|$; hence, the inverse exponential map is the "equal-distance projection" at the identity for S^3.

If the image points in $T_1 S^3$ are denoted w_i^*, then Prentice fits a smooth path to each of the three components of the w_i^* using cubic smoothing splines (see Silverman [12]), and hence obtains a smooth curve in $T_1 S^3$. Prentice then transfers this smooth curve back to S^3 using the exponential map followed by the inverse rotation of S^3.

A serious drawback to Prentice's procedure is the distortion caused by his method of projecting the data points from S^3 to $T_1 S^3$, i.e., the fact that he uses a single projection to handle all of the data, whether they are close to the pole 1 or far away. A second difficulty (which Prentice discusses) is the fact that Prentice's procedure assumes that the error structures at the points w_i^* are isotropic. There is ample evidence that in practice these error structures will not be isotropic (see, for example, Stock and Molnar [13] and Molnar and Stock [8]). Prentice points out that his procedure can be modified to handle the case that the error distributions at the w_i^* are not isotropic but have the same fixed covariance matrix; however, this condition will not usually be satisfied in practice.

An important article related to the fitting problem discussed above is that of Jupp and Kent [7]. Jupp and Kent consider a similar fitting problem in one lower dimension. Let (τ_i, w_i), $i = 1, 2, ..., n$, be a sequence of time-labeled points where the τ_i are strictly increasing and $w_i \in S^2$, the unit

sphere in $\mathbb{R}^3$. Jupp and Kent consider the problem of fitting a smooth curve $\hat{g}(t)$ to this data, where $\hat{g}(t)$ takes values in S^2. In order to transfer this fitting problem to a Euclidean space, Jupp and Kent use a method less prone to distortion than the method of projecting onto a fixed tangent space, namely, they use the method of "unrolling" a path $g(t)$ on S^2 onto a path $g^*(t)$ in $T_k S^2$, where k denotes the north pole on S^2, i.e., $k = (0, 0, 1)^t$. In addition, they use the method of equal-distance projection to project a neighborhood of $g(t)$ on S^2 onto a neighborhood of $g^*(t)$ in $T_k S^2$. Thus, a data point $w_i \in S^2$ corresponding to time τ_i is "unwrapped" from S^2, relative to a base path $g(t)$ on S^2, onto a point $w_i^* \in T_k S^2$. (Note: Jupp and Kent's unrolling procedure is discussed more extensively in Section 3 below.)

Jupp and Kent define a "spherical smoothing spline" on S^2 as follows. First, given a data set

$$D^* = \{(\tau_i, w_i^*) : i = 1, ..., n\},$$

where $w_i^* \in \mathbb{R}^2$, they define the objective functional

$$H_\lambda^*(g^*; D^*) = \sum |g^*(\tau_i) - w_i^*|^2 + \lambda \int_T |(g^*)''(t)|^2 \, dt,$$

where $T = [\tau_1, \tau_n]$ and λ is a positive smoothing parameter. Next, given a data set

$$D = \{(\tau_i, w_i) : i = 1, 2, ..., n\},$$

where $w_i \in S^2$ and given a path $g(t)$ on S^2, the set of unwrapped data points in $T_k S^2$ relative to $g(t)$ is denoted $D^*(g)$. The "spherical smoothing spline with parameter λ for the data set D" is the function $\hat{g}$ such that, when the data are unwrapped relative to $\hat{g}$, $\hat{g}^*$ minimizes the objective function

$$H_\lambda^*(g^*; D^*(\hat{g}))$$

with respect to g^*.

Jupp and Kent present a method to construct a discrete approximation to $\hat{g}$ by an iterative process and report that in practice this iteration seems to converge quickly. Briefly, the iterative process is as follows. First, one subdivides the time intervals $[\tau_i, \tau_{i+1}]$, obtaining a fine grid of points of mesh δ, where δ is a given small number. An initial approximation $g^{(0)}$ is the piecewise great circle path on S^2 which travels from w_i at time τ_i to w_{i+1} at time τ_{i+1} at constant speed, for each i. In general, the ℓ-th approximation $g^{(\ell)}$ is a piecewise great circle path between the mesh points. The $(\ell + 1)$-st approximation $g^{(\ell+1)}$ is obtained from $g^{(\ell)}$ by first unrolling

$g^{(\ell)}$ onto the plane, getting an unrolled curve $g^{(\ell)*}$, and unwrapping the data points relative to $g^{(\ell)}$ to get points $(\tau_i, w_i^{*(\ell)})$. A smoothing spline is fitted in the plane to the unwrapped points and then wrapped back onto S^2, relative to $g^{(\ell)*}$; the new curve on S^2 is $g^{(\ell+1)}$.

3. PARALLEL TRANSLATION VS LEFT- AND RIGHT-TRANSLATION ON S^3

As Jupp and Kent mention in their article, the techniques which they propose can be extended to more general Riemannian manifolds. For example, in the process of "unrolling" a path $g(t)$ on S^2 onto a path $g^*(t)$ in $T_k S^2$, the unrolled curve $g^*(t)$ satisfies

$$(g^*)'(t) = R(t)\, g'(t),$$

where $R(t)$ is a certain rotation of $\mathbb{R}^3$. The rotation $R(t)^t$ corresponds to parallel translation along the curve $g(t)$. Namely, if $v \in T_k S^2$ and $w(t) = R(t)^t v$, then $w(t)$ is a parallel vector field along $g(t)$. Similarly, as Jupp and Kent mention, the technique of "equal-distance projection" corresponds to the method of geodesic normal coordinates in differential geometry. Hence, Jupp and Kent's method can be extended to fitting problems on S^3.

To be specific, let $g(t)$ be a path on S^3, $\tau_1 \leqq t \leqq \tau_n$. For convenience, assume $g(\tau_1) = 1$. Define the rotation $R(t)$ of $\mathbb{R}^4$ by

$$\begin{aligned} R(t)^t\, R'(t) &= M(t), \qquad \tau_1 \leqq t \\ R(\tau_1) &= I, \end{aligned} \tag{3.1}$$

where $M(t) = g(t)\, g'(t)^t - g'(t)\, g(t)^t$. Then, as in the S^2 case, $R(t)^t$ is the length-preserving linear mapping of $T_1 S^3$ onto $T_{g(t)} S^3$ defined by parallel transport along g. We can define the unrolling of $g(t)$ onto $T_1 S^3$ to be the curve $g^*(t)$ determined by

$$\begin{aligned} (g^*)'(t) &= R(t)\, g'(t), \qquad \tau_1 \leqq t \\ g^*(\tau_1) &= 0. \end{aligned}$$

However, in the case of $S^3 = \mathbb{Q}$, there are other methods of unrolling which are connected directly to the group structure of $\mathbb{Q}$, namely, unrolling by left-translation and unrolling by right-translation. To describe unrolling

by left-translation, let $g(t)$ be a path on S^3, $\tau_1 \leqq t \leqq \tau_n$. Define the rotation $R_L(t)$ of $\mathbb{R}^4$ by the equation

$$R_L(t)\, x = \overline{g(t)}\, x \qquad \text{for all} \quad x \in \mathbb{R}^4.$$

(Here matrix multiplication is used on the left; the multiplication on the right is quaternionic. For the fact that $R_L(t)$ is a rotation of $\mathbb{R}^4$, see, for example, Porteous [9, p. 182].) Note that $R_L(t)$ maps $g(t)$ onto 1 and $T_{g(t)}S^3$ onto $T_1 S^3$. The unrolling $g^*(t)$ of $g(t)$ onto $T_1 S^3$ is defined by

$$(g^*)'(t) = R_L(t)\, g'(t) = \overline{g(t)}\, g'(t), \qquad \tau_1 \leqq t$$
$$g^*(\tau_1) = 0.$$

Unrolling by right-translation is similar, except that $R_L(t)$ is replaced by $R_R(t)$ where

$$R_R(t)\, x = x\overline{g(t)} \qquad \text{for} \quad x \in \mathbb{R}^4.$$

To decide which of parallel-, left-, or right-translation is most natural, we first note that splining is fundamentally tied in with our understanding of the error structure connected to the w_i's as estimates of the $g(\tau_i)$'s. To see this, consider the case of splining in Euclidean space, where the w_i^*'s are estimates of the $g^*(\tau_i)$'s. The smoothing spline $g^*(t)$ is gotten by minimizing the objective functional $H_\lambda^*(g^*; D^*)$ defined in (4.1) in Section 4. The first term of this functional essentially measures the size of the deviations $w_i^* - g^*(\tau_i)$.

We now discuss how the errors in tectonic plate reconstruction are understood.

When two plates are diverging at a mid-oceanic ridge, oceanic crust extrudes at the ridge and is carried, by the two plates, away from the ridge as if on conveyor belts. If we were to fix a certain age (say 20 million years) and were able to locate the crust with that age, we would have two *isochrons*, one on each plate. Because the crust hardens immediately after extrusion, the two isochrons have the same shape, that is the shape of the ridge at the time of extrusion of the isochrons.

It follows that there is an $A \in \mathrm{SO}(3)$ such that if (u, v) are points on the two isochrons which extruded from the same point on the ridge, then $v = Au$. Remembering that the isochrons were coincident at their time of creation, we see that the matrix A describes the relative motion of the V-plate in a coordinate system fixed in the U-plate. Alternatively, A^t describes the relative motion of the U-plate in a coordinate system fixed in the V-plate. Note that both the U-plate and the V-plate move, so A and A^t are

relative plate motions, not motions relative to some universal fixed coordinate system.

In a pioneering series of papers on the errors in tectonic plate reconstructions, Stock and Molnar [8, 13] established heuristically a connection between the error structure of the rotation $A^t\hat{A}$ and the geometry of the isochrons. Referring to Fig. 1, we see that $\hat{A}^t v$ represents the best fit back-reconstruction of the V-plate to the U-plate. (The isochrons in Fig. 1 are shown in typical staircase shape. The segments of the isochrons which are parallel to the general trend of the ridge we call "magnetic anomalies"; the segments which are roughly perpendicular to this trend are called "fracture zones.") A fundamental principle of statistics is that the rotations A which are considered consistent with the data (that is, are in the 95% confidence region) are those which do not significantly degrade this best fit reconstruction. These can be obtained by further rotating $\hat{A}^t v$ by a small rotation (that is, a rotation with a small angle).

Mathematically, let $h = (h_1, h_2, h_3)^t$ and define

$$M(h) \equiv \begin{pmatrix} 0 & -h_3 & h_2 \\ h_3 & 0 & -h_1 \\ -h_2 & h_1 & 0 \end{pmatrix}.$$

The matrix exponential $\exp M(h)$ is the right-hand rotation of $|h|$ radians around the axis $h/|h|$. (We use "exp" to denote both the map from $T_1 S^3$ to S^3 described in Section 2 and the matrix exponential. The meaning will be clear from the context.) Reasonable reconstructions $A^t v$ are of the form $\exp M(h)\, \hat{A}^t v$ where $|h|$ is small. In other words, $\hat{A} = A \exp M(h)$ where h is small. Stock and Molnar's insight was to relate the small rotation $\exp M(h)$

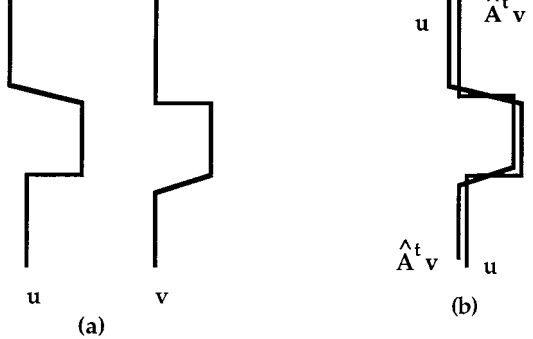

FIG. 1. (a) Corresponding isochrons on the U- and V-plates. (b) Best fit back-reconstruction of the V-plate to the U-plate.

to the geometry of the plate boundary expressed in a coordinate system fixed in the U-plate. In particular, they showed geometrically that if $h/|h|$ is near the center of the data, a small rotation of $|h|$ radians degrades the fit much less than if $h/|h|$ is perpendicular to the center of the data. (See Fig. 2. In this figure the radius of the earth is taken to be 1, d denotes the length of the portion of the plate boundary under consideration, and we assume $d \ll 1$. For convenience, we let $\theta = |h|$.)

Chang [1, 2] mathematized the Stock and Molnar result by relating the spectral decomposition of the covariance matrix Σ^* of h to the geometry of the isochron in its present day configuration on the U-plate.

In terms of quaternions, if $\hat{q}$ and q are the quaternionic representations of $\hat{A}$ and A respectively, then $\bar{q}\hat{q} = \exp(h)$. Under several reasonable

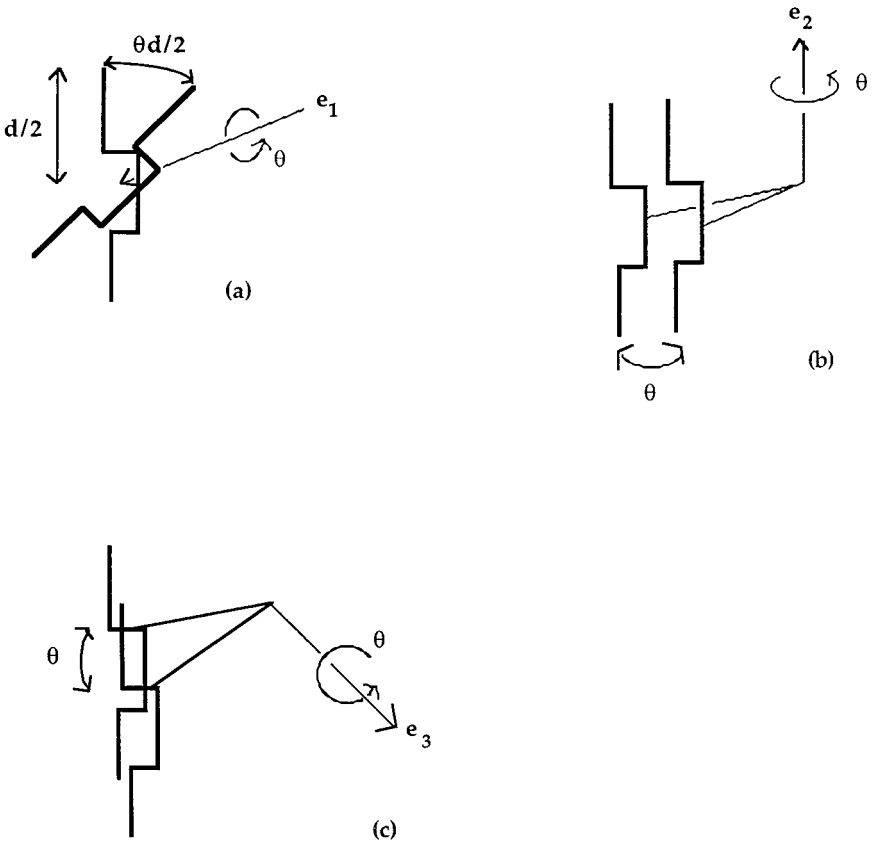

FIG. 2. (a) Rotation around axis e_1 in the center of data skews the isochrons by a maximum of $\theta d/2$. (b) Rotation around axis e_2 parallel to magnetic anomalies results in a mismatch of the anomalies by about θ. (c) Rotation around axis e_3 perpendicular to e_1 and e_2 results in a mismatch of the fracture zones by about θ.

models, the asymptotic distribution of h is multivariate normal with mean 0 and it follows that an asymptotic 95% confidence region for q is

$$\mathscr{C} = \{\hat{q}\exp(h) : h^t(\Sigma^*)^{-1}h < \chi^2_{3,.95}\}.$$

Let $\lambda^{(\ell)} > 0$ and $q^{*(\ell)}$ $(\ell = 1, 2, 3)$ be the eigenvalues and eigenvectors of $(\Sigma^*)^{-1}$ and let $\tilde{Q}$ be the 4×4 matrix with eigenvalue zero corresponding to the eigenvector $\hat{q}$ and eigenvalues $\lambda^{(\ell)}/\chi^2_{3,.95}$ corresponding to the eigenvectors $\hat{q}q^{*(\ell)}$. Since $\exp(h) = 1 + h + O(|h|^2)$, the confidence region $\mathscr{C}$ is asymptotically equivalent to

$$q^t\tilde{Q}q < 1.$$

For the data times τ_i, the 4×4 matrices $\tilde{Q}_i$ referred to in Section 1 are derived in this manner. We have $w_i = \hat{q}_i$ and an asymptotic 95% confidence region for q_i is

$$\mathscr{C}_i = \{\hat{q}_i\exp(h) : h^t(\Sigma_i^*)^{-1}h < \chi^2_{3,.95}\}. \tag{3.2}$$

If we let $\lambda_i^{(\ell)}$ and $q_i^{*(\ell)}$ be the eigenvalues and eigenvectors of $(\Sigma_i^*)^{-1}$, then the positive eigenvalues of $\tilde{Q}_i$ are $\lambda_i^{(\ell)}/\chi^2_{3,.95}$ and the corresponding eigenvectors are

$$q_i^{(\ell)} = w_i q_i^{*(\ell)}.$$

Now consider how the data points w_i and their associated error distributions are unwrapped to T_1S^3, depending on whether parallel-translation, left-translation, or right-translation is used. In the case of left-translation, the point w_i is unwrapped to a point $w_i^* \in T_1S^3$ by the equation

$$w_i^* = g^*(\tau_i) + h_i,$$

where $\exp(h_i) = \overline{g(\tau_i)}\, w_i$. The eigenvectors $q_i^{(\ell)}$ of $\tilde{Q}_i$ are mapped into vectors in T_1S^3 by left-multiplication by $\overline{w_i}$; hence, the $q_i^{(\ell)}$ are mapped onto the eigenvectors $q_i^{*(\ell)}$ of the matrix $(\Sigma_i^*)^{-1}$ defining the confidence region (3.2). In this case we see that the deviation $w_i^* - g^*(\tau_i)$ depends only on $g(\tau_i)$. The covariance matrix of the unwrapped error distribution is independent of the curve g and is related to the geometry of the plates in a coordinate system fixed in the U-plate.

If parallel translation is used, the unwrapped data point w_i^* is given by

$$w_i^* = g^*(\tau_i) + h_i,$$

where $\exp(h_i) = R(\tau_i)\, w_i$. The eigenvectors $q_i^{(\ell)}$ are mapped into vectors in T_1S^3 by, first, mapping them from $T_{w_i}S^3$ to $T_{g(\tau_i)}S^3$ using parallel translation along the geodesic joining w_i and $g(\tau_i)$ and, second, mapping the

resulting vectors to $T_1 S^3$ by means of $R(\tau_i)$. Note that the deviation $w_i^* - g^*(\tau_i)$ and the covariance matrix of the unwrapped error distribution depend on the entire curve $g(t)$ for $\tau_1 \leqq t \leqq \tau_i$. Besides being somewhat unnatural, this leads to a more mathematically complex splining algorithm. For this reason, we prefer unrolling by left group multiplication.

To compare left-translation and right-translation, notice that if $A^t \hat{A} = \exp M(h)$, then $\hat{A} A^t = \exp M(Ah)$. The covariance matrix of Ah is $A\Sigma^* A^t$. The eigendecomposition of $A\Sigma^* A^t$ has the same relationship to the geometry of the (present day) location of the isochron on the V-plate as Σ^* has to the geometry of the isochron on the U-plate. In essence, if A is defined by $v = Au$, but we unroll by right-translation, we are using coordinate systems fixed in the V-plate to express the errors in $\hat{A}$. Notice, however, that A itself is the rotation of the V-plate in a coordinate system fixed in the U-plate. This makes unrolling by left-group translation preferable to unrolling by right-group translation.

Finally, we note that if a rigid body moves in Euclidean 3-space, the motion $y(t)$ of any point on the body can be described by $y(t) = A(t)\, y(0) + b(t)$ where $A(t) \in \mathrm{SO}(3)$. If $g(t)$ is the quaternionic representation of $A(t)$, then $R_R(t)\, g'(t)$ is the angular velocity of the body in "space" coordinates, that is, a fixed universal coordinate system. $R_L(t)\, g'(t)$ is the angular velocity in "body" coordinates, that is, the re-expression of $R_R(t)\, g'(t)$ in a coordinate system which is fixed relative to the body. The *Euler equations*, which are the rotational version of $F = ma$, relate the derivative of the angular velocity in body coordinates to the torques and the moment of inertia tensor. Note that the moment of inertia tensor, which is determined from the distribution of mass throughout the rigid body, is constant when expressed in body coordinates. In view of the form of the smoothness penalty term in $H_\lambda^*(g^*; D^*)$, unrolling by left-translation would seem to be the natural choice for this context.

4. FITTING A SMOOTH HISTORY

First, consider a collection of time-labeled data points in $\mathbb{R}^3$:

$$D^* = \{(\tau_i, w_i^*, \Sigma_i^*) : i = 1, 2, ..., n\}.$$

Here $\tau_1 < \tau_2 < \cdots < \tau_n$, $w_i^* \in \mathbb{R}^3$, and Σ_i^* is the covariance matrix for the distribution from which w_i^* is drawn. We define the objective functional

$$H_\lambda^*(g^*; D^*) = \sum (w_i^* - g^*(\tau_i))^t (\Sigma_i^*)^{-1} (w_i^* - g^*(\tau_i)) + \lambda \int_{\tau_1}^{\tau_n} |(g^*)''(t)|^2 \, dt \tag{4.1}$$

for C^2 functions g^* taking values in $\mathbb{R}^3$. Here λ is a positive smoothing parameter and $|\cdot|$ denotes the Euclidean norm. The function g^* which minimizes (4.1) we call the "smoothing spline with parameter λ for the data D^*."

It is well-known that the components of the minimizing function g^* are natural cubic splines with knots at the points $\{\tau_i\}$. The parameter λ controls the amount of smoothing. As $\lambda \to 0$ the smoothing spline approaches the spline which exactly interpolates the data points; as $\lambda \to +\infty$ the smoothing spline approaches the linear function g^* which minimizes the first term in the expression (4.1).

Note that in general the principal directions of the covariance matrices Σ_i^* will vary with i. Hence, minimization of the objective functional $H_\lambda^*(g^*; D^*)$ cannot be reduced to separate minimization problems involving scalar-valued functions, but must be undertaken as a whole. The algorithm which we use to minimize $H_\lambda^*(g^*; D^*)$ follows the procedure outlined in Silverman [12, Section 6], adapted to our situation.

Now consider a collection of time-labeled data points on S^3:

$$D = \{(\tau_i, w_i, \tilde{Q}_i) : i = 1, 2, ..., n\}.$$

Here the τ_i's are strictly increasing, as before, $w_i \in S^3$, and $\tilde{Q}_i$ is as described in Sections 1 and 3 above. Given a base path $g(t)$ on S^3 the data set D can be unwrapped to a data set $D^*(g)$ in $T_1 S^3$ by means of left-translation.

We seek a "smoothing spline" $g(t)$ on S^3 which in some sense smoothes our original data set D. In analogy to the procedure of Jupp and Kent, we define the "smoothing spline on S^3 with parameter λ for the data D" to be the curve $\hat{g}(t)$, taking values in S^3, such that when the data D are unwrapped to $T_1 S^3$ relative to $\hat{g}$, the unrolled curve $\hat{g}^*$ is the smoothing spline in $T_1 S^3$ with parameter λ for the data $D^*(\hat{g})$. We calculate $\hat{g}$ by an iterative procedure similar to that used by Jupp and Kent.

We do not have proofs of existence or uniqueness for the function $\hat{g}$, nor do we have a proof of convergence for the iterative procedure. Our situation in these respects is similar to Jupp and Kent's. However, as with Jupp and Kent, we have found that in the examples we have considered the iterative procedure converges quickly, usually within two or three iterations.

5. CHOOSING THE VALUE OF THE SMOOTHING PARAMETER

A method often used for determining the value of the smoothing parameter λ, when doing spline smoothing, is cross-validation. For a given

TABLE I

Central Atlantic Data

Time (m.y. before present)	Estimated rotation Axis latitude	Axis longitude	Rotation angle
10.0	77.55°N	85.90°E	2.40°
20.0	80.99°N	35.90°E	5.29°
35.5	75.60°N	3.02°E	9.95°
49.5	75.21°N	2.27°W	15.11°
59.0	80.93°N	1.01°W	18.09°
67.5	83.53°N	4.30°E	20.77°
72.5	82.35°N	10.38°W	22.75°
74.3	82.45°N	13.06°W	23.68°
80.2	78.89°N	17.94°W	26.93°
126.0	66.25°N	19.40°W	56.39°
131.5	65.87°N	18.64°W	57.52°
149.5	66.27°N	18.08°W	62.06°

value of λ, a "cross-validation score" is determined as follows: one leaves out each data point, one at a time, and measures how closely the spline fitted to the remaining points fits the point left out; the cross-validation score is the sum of squares of the norms of the deviations found this way. The optimum value of λ is then taken to be the value which minimizes the cross-validation score. (See, for example, Silverman [12, p. 5] or Wahba [14, p. 47].)

In our case, suppose we have data $(\tau_i, w_i^*, \Sigma_i^*)$, $i = 1, 2, ..., n$, where $w_i^* \in \mathbb{R}^3$ and the Σ_i^* are the corresponding covariance matrices. For a fixed value of i, let $g_\lambda^{-i}(t)$ denote the smoothing spline calculated using the smoothing parameter λ and all data points except the ith one. It is natural to define the cross-validation score $\mathrm{XVSC}(\lambda)$ by

$$\mathrm{XVSC}(\lambda) = \sum_{i=1}^{n} (w_i^* - g_\lambda^{-i}(\tau_i))^t (\Sigma_i^*)^{-1} (w_i^* - g_\lambda^{-i}(\tau_i)).$$

However, our experience, based on a Central Atlantic data set kindly made available to us by Joanne Stock of the California Institute of Technology, is that the method of cross-validation leads to over-smoothing. This data set comprises $n = 12$ points. In Table I we list the times τ_i and estimated rotations $\hat{A}_i$. For reasons of space we omit the matrices $\tilde{Q}_i$ which define the 95% confidence regions. The method of cross-validation applied to this data set gives the optimal value of λ:

$$\lambda_{\mathrm{opt}} = .295 \times 10^{10}.$$

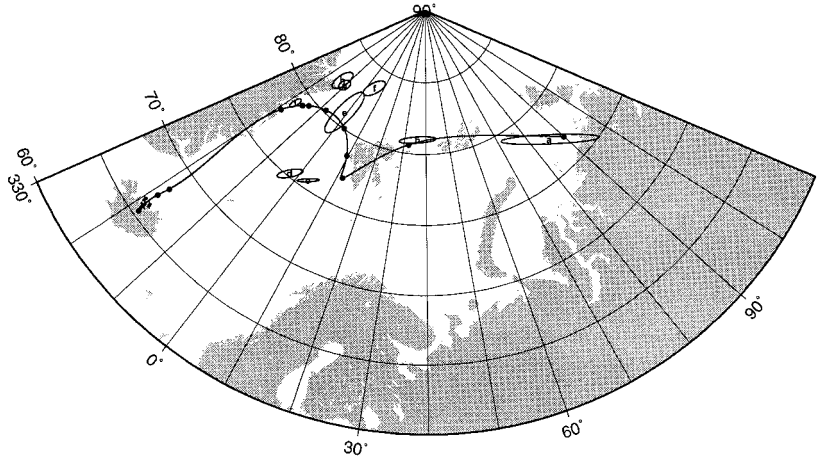

FIG. 3. Fitted path of the rotation axis for the Central Atlantic data when the smoothing parameter is determined by cross-validation.

However, the corresponding fitted spline on S^3 misses the given 95% confidence regions badly. For only one value of i (out of twelve) does the fitted rotation $\hat{A}(\tau_i)$ lie in the corresponding confidence region. The effect of the over-smoothing can be seen graphically in Fig. 3. In this figure the small letters "a", "b",..., indicate the estimated axes of rotation at the data times. (Point "a" corresponds to 10 *m.y.* before present, point "b" corresponds to 20 *m.y.* before present, etc.) Each estimated axis is accompanied by the region of acceptable rotation axes corresponding to the 95% confidence region for the true rotation. The solid line indicates the fitted path of the rotation axis (with the value of λ determined by cross-validation); the black dots on this line indicate the fitted rotation axes at the data times. From Fig. 3 it is clear that the rotation path obtained using cross-validation gives a poor fit to the data.

We have found that an alternative method of choosing the value of λ does better, namely the discrepancy method. Wahba [14, p. 63] discusses this method in the single-variable case in the case where all the variances are equal. The corresponding method in our case can be described as follows. For a fixed value of λ, let $g_\lambda(t)$ be the smoothing spline in $\mathbb{R}^3$ to the data $(\tau_i, w_i^*, \Sigma_i^*)$, $i = 1, 2, ..., n$. We let $Q(\lambda)$ denote the quantity

$$Q(\lambda) = \sum (w_i^* - g_\lambda(\tau_i))^t (\Sigma_i^*)^{-1} (w_i^* - g_\lambda(\tau_i)). \tag{5.1}$$

The quantity $Q(\lambda)$ is zero when $\lambda = 0$ and will tend to increase as λ increases. We define the optimal value of λ by the equation

$$Q(\lambda_{\text{opt}}) = 3n.$$

The discrepancy method above can be justified heuristically by the following fact. Suppose the $\{w_i^*\}$ are independent random vectors, such that w_i^* has expected value μ_i and covariance matrix Σ_i^*. Then the expected value of

$$\sum_{i=1}^{n} (w_i^* - \mu_i)^t (\Sigma_i^*)^{-1} (w_i^* - \mu_i)$$

is $3n$.

The discrepancy method applied to the Central Atlantic data set in Table I gives the optimal value of λ:

$$\lambda_{\text{opt}} = .62 \times 10^8.$$

The resulting fitted spline on S^3 does much better than that corresponding to cross-validation. For eleven out of twelve values of i, the fitted rotation $\hat{A}(\tau_i)$ lies in the 95% confidence region corresponding to τ_i. Figure 4 illustrates the behavior of the fitted curve determined by the discrepancy method; Fig. 4 is similar to Fig. 3 except that λ is chosen by the discrepancy method rather than by cross-validation. (The one fitted rotation which does not lie in the corresponding confidence region is that for 74.3 *m.y.* before present. This fact is not apparent from Fig. 4 since it displays only the behavior of the fitted rotation axis, i.e., it does not show the behavior of the rotation angle.)

It is perhaps not surprising that the discrepancy method gives better results in our situation than the cross-validation method, as the discrepancy method uses the fact that the variances of the error distributions

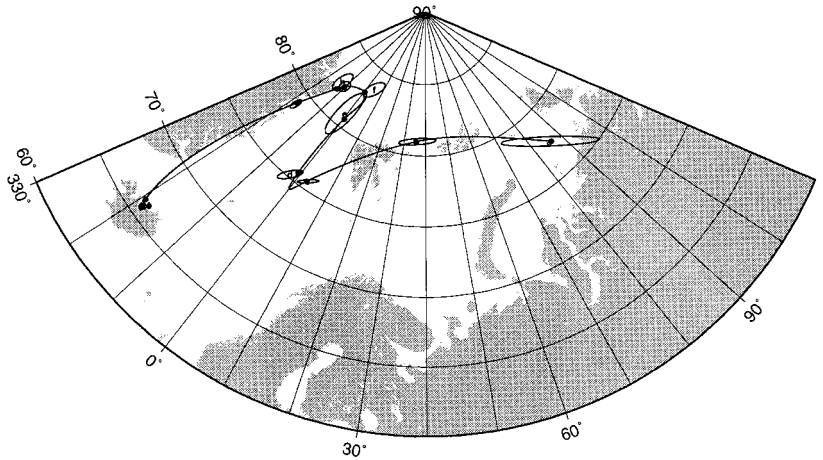

FIG. 4. Fitted path of the rotation axis for the Central Atlantic data when the smoothing parameter is determined by the discrepancy method.

for the data points are known, whereas the cross-validation method requires only knowledge of the relative variances, i.e., the ratios of the variances at the data points to the average variance.

REFERENCES

1. T. Chang, Spherical regression, *Ann. Statist.* **14** (1986), 907–924.
2. T. Chang, On the statistical properties of estimated rotations, *J. Geophys. Res.* **92B** (1987), 6319–6329.
3. T. Chang, On reconstructing tectonic plate motion from ship track crossings, *J. Amer. Statist. Assoc.* **83** (1988), 1178–1183.
4. T. Chang, Spherical regression with errors in variables, *Ann. Statist.* **17** (1989), 293–306.
5. T. Chang, J. Stock, and P. Molnar, The rotation group in plate tectonics and the representation of uncertainties of plate reconstructions, *Geophys. J. Internat.* **101** (1990), 649–661.
6. M. S. Hanna and T. Chang, On graphically representing the confidence region for an unknown rotation in three dimensions, *Comput. Geosc.* **16** (1990), 163–194.
7. P. E. Jupp and J. T. Kent, Fitting smooth paths to spherical data, *Appl. Statist.* **36** (1987), 34–46.
8. P. Molnar and J. Stock, A method for bounding uncertainties in combined plate reconstructions, *J. Geophys. Res.* **90** (1985), 12537–12544.
9. I. R. Porteous, "Topological Geometry," 2nd ed., Cambridge Univ. Press, Cambridge, UK, 1969.
10. M. J. Prentice, Fitting smooth paths to rotation data, *Appl. Statist.* **36** (1987), 325–331.
11. L. Rivest, Spherical regression for concentrated Fisher–von Mises distributions, *Ann. Statist.* **17** (1989), 307–317.
12. B. W. Silverman, Some aspects of the spline smoothing approach to non-parametric regression curve fitting, *J. Roy. Statist. Soc. Ser. B* **47** (1985), 1–52.
13. J. Stock and P. Molnar, Some geometrical aspects of uncertainties in combined plate reconstructions, *Geology* **11** (1983), 697–701.
14. G. Wahba, "Spline Models for Observational Data," Soc. for Industr. & Appl. Math., Philadelphia, 1990.